Hindawi
Security and Communication Networks
Volume 2018, Article ID 4231326, 12 pages
https://doi.org/10.1155/2018/4231326

Research Article

Uncovering Tor: An Examination of the Network Structure

Bryan Monk, Julianna Mitchell, Richard Frank, and Garth Davies

International CyberCrime Research Center (ICCRC), School of Criminology, Simon Fraser University, Burnaby, BC, Canada

Correspondence should be addressed to Julianna Mitchell; jdm17@sfu.ca

Received 3 November 2017; Revised 21 March 2018; Accepted 2 April 2018; Published 9 May 2018

Academic Editor: Zheng Yan

The dark web is a concealed portion of the Internet that can only be accessed through specialized software. Although multiple dark web technologies exist, with a common trait of using encryption to enforce anonymity, the Tor network remains the most prominent dark web network. To visit websites on the network, the user must use a heavily modified Firefox browser. The use of encryption to achieve anonymity poses a significant challenge for law enforcement that wishes to monitor users and content for illicit activity. This study examines Tor by focusing on the network structures created between websites via hyperlinks. Examining hyperlinks can provide insight into how virtual communities form on a network. We explore traditional social disorganization principles as a basis to draw comparisons between these virtual communities and real-life crime-prone neighborhoods. Automated data collection techniques were used to leverage the interconnected nature of domains on Tor. Using social network analysis, website hyperlinks are examined and core sites are identified. The analysis shows that these core sites form a significant portion of all connections made on the network with a density of 0.132. This core serves a critical function and has implications for detecting how users connect on Tor.

1. Introduction

The Internet has been shown to be a very powerful communication tool, enabling individuals to connect globally to access and exchange material with very few obstacles, including governmental or jurisdictional interference. While the Internet has been a significant driving force in the globalization of knowledge, it has simultaneously created an environment where nefarious users can disseminate illicit content, connect users with similar interests, and facilitate the proliferation of virtual illicit communities. Today, the majority of global Internet users still only access a fraction of the Internet, the portion known as "the surface web" [1]. As opposed to the surface web, which is open and reachable by anyone with an Internet connection, the "dark web" uses the infrastructure of the surface web but, through encryption, creates a subnetwork that is anonymous, concealed, largely unindexed, and only accessible through encrypted Internet browsers [1, 2]. A true dark web has three central characteristics [2]: (i) it uses peer-to-peer technology rather than centralized servers that information can be tracked back to; (ii) it uses the infrastructure of the Internet; and (iii) it operates through nonstandard protocols and ports. As a result of these characteristics, the structure of the dark web is constantly evolving. At the same time, these characteristics can foster a potentially risky environment within which illicit behaviors can more easily occur.

Dark webs are employed for a multitude of objectives, including keeping Internet activities and identities anonymous, evading censorship, and communicating sensitive information securely. Simultaneously, dark webs also increase opportunities and support for individuals conducting illicit activities [3]. Online connections can be created instantaneously and those interested in pursuing malicious activities or accessing illicit content can network with little effort [4]. This has led to growing concerns over the extent to which dark webs may be facilitating and fostering serious criminal activity, as well as assisting the operations of terrorists and violent extremist groups [5]. Law enforcement and researchers have realized the need to understand the extent of criminal activity on dark webs but have faced three related challenges. First, the extent of criminal activity remains unclear due to the sheer size and dynamic structure of dark web networks. Second, research is hindered by a lack of adequate tools and methods to examine the immensity and evolving structure of dark webs. Finally, criminological

theories have yet to be applied to the dark web as a means of understanding how crime can flourish in this environment.

This study responds to these challenges by examining dark web-based networks through the use of advanced data collection tools and analytic methods. While many dark web browsers exist, the Tor network remains one of the most well-known and frequently operated networks used to access and operate on the dark web [1] and, as such, is well suited to provide an in-depth understanding of network structures. Tor is a peer-to-peer network that operates by connecting users through specialized software designed to anonymize and encrypt data sent between those users. The Tor network allows users to access the dark web through a specially developed web browser. Since its inception in 2002, Tor has become one of the most universally used dark web technologies due to its anonymity features and its ability to efficiently access dark web content and information [6]. Although Tor developers maintain that the network's primary mission is to provide users with the technology to evade intrusive surveillance and data collection by governments and corporations while simultaneously fostering innovative research in online anonymity and privacy, there are increasing concerns about the illicit opportunities that could be provided to malicious users.

1.1. Content on Tor. A key research interest has been the nature of the content available on the Tor network. Content analyses have revealed that an extremely high volume of unethical content is available through the network [5], where unethical content was defined as negative behavior and included "anti-social behavior, drugs, weapons, hacking, cannibalism, bomb making, hit man services, black markets, and child pornography" [5]. Some researchers have concluded that the prominence and pervasiveness of this unethical content far outweigh the benefits of Tor services [5]. However, such findings have been based exclusively on content found on forum discussions from three main databases, potentially limiting their generalizability [5]. Past research has also demonstrated a varying degree of illegal activity on the network where virtual interaction and illegal file sharing were found to be the most prominent illegitimate uses of dark webs such as Tor [7]. File sharing includes sharing copyright-infringing files, such as music, movies, and TV [5]. Cybercriminals also have been known to utilize the network in order to exchange information and transfer data for hacking, identity fraud, and buying and selling illegal goods [8].

1.2. Malicious Uses of Tor. Tor and similar dark webs allow users to access websites that sell illegal goods and services, similar in functionality and structure to eBay, although carrying illicit material.. Such sites, known as dark markets, carry a variety of illegal commodities and appear to cater to users looking for specific materials. Most of these markets require Bitcoin as virtual trading currency. Bitcoin is an increasingly popular peer-to-peer decentralized payment system [9]. Various types of malware have also been identified on the Tor network. Such malware presents a highly dangerous threat to individuals and uses ransomware to encrypt individual's files and prohibit access to these files until a payment is provided [2]. It has been speculated that terrorist organizations operate on these markets to finance their activities through the selling of illegal weapons and drugs. This is possible for two important reasons: first, unlike the regular Internet, these domains are not registered to a central authority; and second, Tor is highly anonymized, which means users feel safe posting this content.

1.3. Tor Hyperlinks. Tor websites, like those found on the regular Internet, do not exist in a vacuum; rather, they are hyperlinked both with themselves and to each other. Without hyperlinks, websites could only be found if the exact URL was known to the user. Hyperlinks thus form the basis for how users traverse this network, connect to domains, explore content, and connect with other users. Social network analysis allows for the analysis of these hyperlinks and illuminates website connectivity. Websites with more incoming and outgoing hyperlinks may be considered to be more popular and more important to the dissemination of information within the Tor network. Network analysis can help to determine not only how information is being distributed and shared among network structures but also the importance of particular websites to the network. Like other Internet structures, there may be a small dense cluster of nodes that form the center of the network and guide the way information travels. This would have significant implications for how the structures of communities develop and function and could improve strategies used by law enforcement agencies to detect and remove illicit users and content.

In this study, we examine Tor's hyperlink connections through social network measures to gain insight into the network structures that form on Tor and how these structures create conditions favorable to deviant or criminal activity. An automated data collection tool, known as The Dark Crawler (TDC), was employed to collect and capture a sample of dark web content. TDC was able to navigate and collect hyperlink information for 1,220 unique dark web websites on the Tor network. This study also explores whether the core principles of social disorganization theory can explain the prevalence of "bad" communities on Tor. Traditional social disorganization theory hypothesizes that community structures characterized by instability, heterogeneity, and weak social ties can foster crime and disorder [10]. We suggest that Tor social network structures, formed through hyperlinks, have a similar nature in that sparse and unstable networks composed of weak connections could potentially foster illicit online activity and content. Although no known studies have yet to determine the time Tor domains are online and offline, the transitivity and instability of Tor have been widely speculated by researchers [11]. Additionally, like crime-prone neighborhoods, networks that form on Tor lack regulation or monitoring by law enforcement and government. Social disorganization theory proposes that the absence of a governing system can limit community members' ability to formally or informally control the behavior of others and increase opportunities for individuals to engage in criminal activities. If connections on Tor are found to be unstable and weak, coupled with a lack of

regulation, it is possible that these characteristics present unique opportunities for users to leverage the network to engage in devious activity similar to that of socially disorganized communities. There is currently a void in research that has specifically examined this link between dark web online communities and offline communities and, as such, the application of this criminological theory is only used as an exploratory lens in the current study. Previous work on Tor has examined individual websites such as The Silk Road, an online black market, [9, 12] or as a larger construct providing a content analysis of the domains found within the network [2, 5]. Using social network analysis, this study aims to examine (a) hyperlink connections and the structure of website communities they form on Tor and (b) how characteristics of these communities could have implications for criminal activity on Tor as understood through the lens of social disorganization theory. From a law enforcement perspective this may have implications for disruption strategies or target selection for policing interventions.

2. Theoretical Background

Past research has examined how Internet communities created through hyperlink connections can be understood as individuals connecting and developing ties, thereby forming small online communities within a large network [4]. These virtual communities are not constrained by geographic barriers and have provided an avenue for users to connect worldwide [13]. This has resulted in the development of large, globally connected social networks that allow users to connect with like-minded individuals, gain social support, and share information and material. Researchers have suggested that communities that form online have a similar nature to offline communities, where complex social networks are formed through similar interests or purposes and are sustained through continuous interactions among users [7, 14]. Further, researchers have proposed that, like the offline world, the Internet also contains "bad neighborhoods" with crime distributions resembling the offline world. These virtual neighborhoods have been considered to have greater concentrations of crime to the extent that they are largely composed of IP addresses hosting illicit content or performing malicious activities. Malicious users can also acquire criminal capital through formed connections by sharing information and material within online criminal communities [4]. As part of this study, we draw inspiration from a major criminological approach to explore whether this could be true of dark web network connections. While parallels can be drawn between regular Internet structures and dark web structures in how they connect websites, they are distinct enough to warrant a separate examination of Tor's hyperlinking structures. That said, it is probable that Tor is used in the same manner by malicious users. As such, the primary assumptions of social disorganization theory are used as basis to understand how the structure of Tor itself, and how users connect on the network, could potentially increase the opportunities for deviant or criminal activity.

2.1. Social Disorganization Theory. Many criminologists have focused on the relationship between urban and community organization and crime [15–17]. Social disorganization theory specifically draws attention to the reciprocal connection between communities and crime and has become one of the most influential models of crime within criminology over time. Central to the social disorganization approach is the idea that community organization and social ties are important mechanisms through which communities can control crime. As such, the theory suggests urban organization can influence patterns of crime. In particular, the research of [10] proposed that four structural factors, instability, heterogeneity, weak social ties, and lack of supervision, can increase the likelihood of crime and delinquency in a community. It is argued that such factors disrupt the social organization in a community, where social organization is measured by the prevalence and interconnectedness of social networks. As such, they theorized that weak social structures decrease the ability of the community to maintain informal social control over members' behavior. Informal control in a community has been known to occur when members can control crime through informal surveillance of the neighborhood and intervene in problematic or suspicious activity. Strong and cohesive social ties within community networks have, therefore, been shown to increase the effectiveness of social control in reducing crime because community members are more willing to engage in monitoring and guardianship behaviors against crime [16, 18]. Thus, the prevalence of community organization or community disorganization are regarded as separate traits that have an influence on crime rates.

Social disorganization theory has generally been used to analyze urban crime geographies where researchers have focused on demographic, economic, social, familial, and urban factors when assessing criminal activity. The three variables most often assessed are those related to poverty, ethnic heterogeneity, stability, and population mobility in a community, as these are viewed as factors that can weaken a community's ability to manage the prevalence of crime [17]. In examining and testing these factors, researchers have found that such characteristics present in a community do appear to have a relationship with increased crime rates. With regard to the role of poverty, studies have found that communities of low socioeconomic status lack money and resource to organize and mobilize a community [19, 20]. In a study by [21] ethnic heterogeneity present among communities was found to be associated with social disorganization, where the degree of ethnic mix and population density to violent crime was analyzed. With regard to population mobility, meaning the rate of incoming and outgoing community members, [17] found that residential instability had a negative effect on network connections in a community, which in turn was related to an increase in crime. Additionally, [22] found that instability, measured by residents living in a community less than 2 years, had a reciprocal relationship with community disorder where the more instability that was present in a community led to increased disorder, and vice versa.

Network density has also been proposed as an influential characteristic by indicating the extent to which individuals

are connected to each other by direct relations [17]. For instance, high network density within a community has been found to increase its ability to control criminal behavior as community members are more apt to monitor and respond to such behavior. Alternatively, low network density and weak social ties can result in the inability to exert social control [17]. Social disorganization theorists have proposed that supervision is an important component within communities [23]. Crime and disorder can develop where there is lack of supervision over activities and these activities are not held in check through social control. A lack of supervision or regulation of activities within a community can lead to higher rates of crime, whereas cohesive communities are able to exert control and intervene in criminal or delinquent activities. Indeed, in studies on teenage delinquency, researchers found that gang developed in communities when teenagers were left unsupervised whereas more cohesive communities were better able to collectively supervise teenagers and control potentially delinquent behaviors [10, 17].

Overall, criminological studies of social disorganization principle have consistently demonstrated how characteristics of a disorganized community, sparse networks, unsupervised groups, instability, and heterogeneity, are important in explaining variation in crime rates across geographies. However, it is noted that these studies measured concepts related to disorganization and are thus subject to measurement error, with the potential that other underlying variables could be influencing crime rates. Despite such limitations, the results of these studies have proved to be consistent with the key principles proposed by social disorganization theory.

Drawing on social disorganization theory, we propose that Tor structures exhibit similar characteristics to offline crime-prone communities. In this study, these characteristics will be adapted in a more general sense through the results of social network analysis and what such measures reveal about the Tor network. Weak connectivity, low network density, and large diversity within Tor could indicate disorganization within these structures. Structural disorganization could be a contributing factor to the presence of illicit activity and material on Tor. Due to its anonymization features, Tor is also inherently void of regulation and guardianship, which is likely to increase or attract malicious users operating within the network. We propose that examining hyperlink structures is an important step to understanding Tor networks, as hyperlinks play a crucial role in facilitating the transmission of information and content, connecting malicious users and websites, and providing opportunities to engage in criminal activities.

2.2. Current Study. Examining the Tor network, how it is deployed, and who comprises the Tor community is an imperative undertaking to better understand the dark web. The current research aims to achieve an in-depth understanding of the network's structure through the use of innovative collection and analytic techniques. Using a specialized web crawler enables efficient collection and the ability to filter a vast amount of information from the Tor network. Subsequent analysis of this information can help determine the extent to which Tor websites are connecting to other websites within the network through hyperlinks. Social network analysis can assess these connections and provide a macro-level understanding of the network structures within Tor. The social network measures serve as a way to analyze how legitimate and illegitimate users are able to navigate and connect on Tor. These connections are implicit in understanding the structures within the network and determining the impact of each website within this dark web. This can generate insights into what content exists within Tor and how the structures of the network may foster online illicit activity. The overall network form will be examined along with social network measures including cohesion, homophily, and core-periphery to provide insight into whether the network structures are characterized by sparse, weak, and heterogeneous social connections between domains.

Networks can take many forms but generally exhibit two dichotomous structures: random or scale-free. Random networks are composed of a disconnected set of nodes that are paired with a certain probability [24]. These networks often have low heterogeneity, in contrast to real world observed networks where edge formation is likely a product of choice [24]. Scale-free networks often follow a power-law distribution where nodes are more likely to connect to a central few nodes rather than forming connections to only unique nodes [25]. New nodes in the network are more likely to select edges to these central actors rather than follow a random pattern. These scale-free networks characterize the regular Internet, which may provide insight into the connections between nodes on the dark web (Tor). Scale-free networks may contain a central core of hubs that contain most of the connections within the network. These hubs then are important to how information and users traverse the network. Identifying these hubs through a core-peripheral model will provide insight into how a user may traverse the network with an additional focus on how criminal activity may proliferate within Tor. Legal domains acting as hubs may provide access and opportunities to illicit ones, which would be largely isolated otherwise. The presence of a core then reduces the distance that a user will have to travel to find illicit content increasing criminal opportunities.

3. Data and Methods

The Dark Crawler (TDC) is a modified version of the webcrawler presented in previous research [26–28] and is shown in detail in Appendix. It operates by collecting and downloading webpages into an offline database. Prior to data collection, a list of seed websites was compiled to provide the crawler points of departure. Seed websites are universal resource locators (URLs) that are manually selected and can be as few as one website or more than 1,000 websites. In the current study, Tor domains were used as seed websites. These Tor domains were found through a Google search of the most widely known.onion directory called the Hidden Wiki, with additional URLs found on Reddit and other regular Internet sites. The Hidden Wiki categorizes and shares known Tor domains, allowing users to search for various types of legal and illicit content. Overall, a collection of 150 seed websites

were identified to start the search process. These 150 websites were selected at random using a random number generator and through content analysis the domain was categorized (Drugs, Child Exploitation, Directory, Hosting, Weapons, etc.) The categories were modeled from previous studies [2, 5] with some categories collapsed into others (guns and hitman services were coded as weapons) while others were expanded. Often the "Other" or "Misc." categories were used as a catchall for hard to identify websites and those were disaggregated for this study into appropriate subcategories such as politics.

Starting with the seed websites, the crawler downloaded each top level domain and added any found hyperlinks containing a.onion address to its internal queue. It recursively followed these links from the queue until either no.onion domains were found or it reached the termination criteria. For the purposes of this study, the termination criterion was the downloading of 1 million webpages across any number of domains. Among the 1 million Tor webpages that comprise the sample there were 1,220 unique domains.

3.1. Social Network Analysis

3.1.1. Homophily. Homophily looks to explain why nodes with similar features are more likely to share social relationships or ties [29]. This relationship has not been explored through the analysis of the links between top level domains in an online network. If domains are more likely to share hyperlinks with similar domains, this can be compared to homophily observed through website content. Network topography may play a more prominent role in which webpages hyperlink to other webpages rather than simply connecting to webpages with similar content. This has implications about the nature of node hyperlinkages operating within the dark web. Tor nodes are not necessarily looking to increase visibility and might only be interested in remaining part of the underground community. The presence of homophily within the network would indicate the possibility that websites were more likely to form connections with similar others. Alternatively, a lack of homophily would indicate that the network structure was not driven by homogeneous website connections but instead was driven by a motley crew of users and content. If network connections are dissimilar and diverse, this could be a contributing factor to the presence of illicit activity on Tor. As social disorganization theory proposes, diversity within a community can lead to higher delinquency as it interferes with members' ability to develop strong social connections and effectively work together to communicate and provide surveillance within the community [10].

3.1.2. Cohesion. A scale-free network features highly centralized nodes or hubs connecting to other nodes forming large clusters [30]. Traditionally, a network that is characterized by the small-world phenomenon has an average path length of 6 [31]. The Internet is characterized as having small-world features despite containing an average path length of 19 [30]. This is due to the sheer size of the Internet which follows logarithmic scaling, where the degree centralization remains relatively high despite the sheer volume of domains present [32]. Similar to the regular Internet, Tor will likely follow these properties, which has important considerations regarding network traversal, information flow, and law enforcement disruption.

Given the novel nature of examining Tor through social network analysis, examining these properties was essential before comparing results and the generalizability of this network to the regular Internet. If Tor does not resemble the regular Internet, then comparing the importance of the nodes within the network would be ill-advised. Determining the nature of the Tor network would also allow additional network measures to be examined on the overall network. A core-periphery analysis was conducted to determine the importance that some hubs might be having on the overall network. If there was no identified core, this would indicate that the websites were connected to each other and had a similar number of network ties. Alternatively, the presence of a core would suggest that the Tor network was operated by a few centralized hubs that serve as crucial links to the entire network. As the adoption of Tor by users remains relatively low, accessing content on Tor can be difficult without prior knowledge of where to look. To remedy this problem, it is likely that central hubs have emerged which are responsible for linking new users to onion domains. Without hubs directing users to various domains, Tor users would have difficulty accessing desired content. The Tor network takes advantage of the high density of users to operate on a peer-to-peer system which increases anonymity for the entire community. Social disorganization theory would suggest that this anonymity within a dense network increases criminal activity as it interferes with accountability to other Tor users. Further, opportunities to engage in illicit activity on Tor are not constrained by isolation. While some studies have suggested that living in a sparsely populated area can reduce opportunities for offending due to the offender's distance from targets [33], this is unlikely to impact deviant users in the online realm. Users wishing to access or share illicit content can do so without being constrained by geographical barriers and likely desire to remain hidden and disconnected on the network to avoid law enforcement.

3.2. Core-Periphery Analysis. Core-periphery analysis has been used successfully on smaller pockets of the Internet. Researchers hypothesize that the Internet, due to its architecture, retains an inherent core structure [34]. Although researchers suggest that the core is formed by connections based upon network traffic, hyperlinks serve as another way to measure connectivity. Most of the research focusing on online communities uses core-periphery to look at user social relationships [35, 36], but few use websites to examine these connections. While the existence of a core is briefly discussed in [10], no social network analysis regarding the structure was conducted. We propose that while a core may be present in the network and even necessary for users to find content, hyperlink connections outside of the core are likely to be infrequent and weak. Tor users interested in accessing or distributing illicit content can be expected to remain isolated in order to avoid detection. Assuming website linkages are

Table 1: Comparison of categorical and continuous core-periphery models.

One mode categorical $N = 61$	One mode continuous *minres* $N = 8$	One-mode continuous score (Corene)	Website content	Website type
N531	N531	0.76	Offline	Offline
N514	N514	0.31	Links	Directory
N1007	N1007	0.26	Wiki	Hosting
N899	N899	0.20	Politics	Directory
N938	N938	0.20	Politics	Directory
N937	N937	0.19	Politics	Directory
N898	N898	0.19	Politics	Directory
N515	N515	0.16	Links	Directory
Fit = 0.22		Fit = 0.07		

a valid form of network construction, and the nature of the Tor network is to be covert, then examining offline covert networks appears to be the logical choice from which to draw comparisons. The Tor network most closely resembles criminal networks, where the actors are attempting to avoid detection. The types of offline networks that share these characteristics would most likely be gang affiliation networks, given that actors are actively trying to avoid detection and are interested in conducting illicit activities.

Core-periphery analysis has been shown to be effective in determining who the central actors are within numerous networks, both offline and online [34, 37, 38]. Core-periphery analysis can potentially reveal websites that provide highly important positions within the network, even though they would not appear to be influential based upon network scores or using the naked eye. A person may appear highly influential in a network while actually not being a central component of the core. Some nodes that may appear as peripheral entities actually have significance to the core of the network and would be overlooked. Previous studies utilizing the various forms of the TDC have discussed the biases and limitations associated with using this data collection technique [26]. The current sample of the Tor network relied upon the manual data collection of seed websites through content analysis which may influence what Tor nodes are represented within the network [26, 28].

4. Results

The analysis of the 1,220 nodes in the Tor sample revealed a sparse network with an average degree score of 2.27. Each node was connected to only a couple of other websites, showing the scale-free nature of the network. There were 2,763 ties between the nodes, leading to a network density of 0.002. This indicated that less than one percent of all possible connections exist within the network. Websites within the network were only connected to a few select nodes. These ties were directed, as not all hyperlinks are reciprocated back between websites. Of the 2,763 edges within the network only 136 were reciprocal between nodes, while 2,627 were not. Node reciprocity then only happened in 4.9% of all connections within the network. The average distance of this network was 4.95, meaning that it took, on average, about 5 connections to get from one end of the network to the other when considering all possible links between pairs of two nodes. The surface web, on average, has an average distance of 16–19 connections [25], suggesting that Tor exhibits a different structure. The average score additionally suggested that the network contained properties associated with small-world networks. The degree centralization for the network is 41.88%, which indicated there were scale-free features present within the network. We suggest the identification of these sparse connections and a lack of reciprocity between the hyperlinks contributes to the presence of crime among the Tor community as compared to the surface web. This allows domains to remain hidden from unwanted attention, such as law enforcement, while still retaining the ability to attract interested users.

4.1. Categorical versus Continuous Models. Multiple tests were conducted using core-periphery analysis to determine the best model fit for the dataset. The final results, reported in Table 1, compared the categorical model and the one-mode continuous model using the *minres* algorithm. The *minres* algorithm is a form of factor analysis that can be used if the diagonal values within the relational matrices are not valued and only detect the presence or absence of a tie [39]. The *minres* algorithm works best with binary networks where the presence of a tie is captured. Model fit is represented as a Pearson's correlation coefficient between the current model and an ideal model with a maximized core structure [39]. The procedure inherently maximized nodes with connections into one block in a matrix (the core) while the nodes with few connections (the periphery) were minimized and placed into a second block [39]. This immediately biased any significance testing and thus model fit was largely a descriptive process, rather than reaching an arbitrary cut-point. The categorical model suggested a core of 61 websites with a periphery of 1,159 nodes and had a model fit of 0.22. The one-mode continuous analysis was conducted using the *minres* algorithm and a core of eight nodes with 1,212 periphery nodes was suggested with a model fit of 0.07. Comparison between the categorical model and the continuous model demonstrated the categorical model was a better fit for the data given the large differences between the scores in model fit. Table 1 contains the top 8 websites as

Table 2: Group densities between core and periphery across categorical and continuous models.

Categorical (Fit = 0.22)	Core	Periphery	Continuous (Fit = 0.07)	Core	Periphery
Core	0.132	0.019	Core	0.250	0.160
Periphery	0.001	0.001	Periphery	0.001	0.001

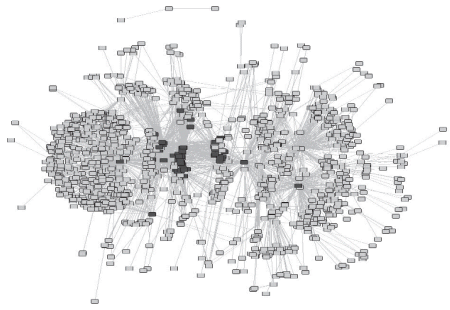

Figure 1: Core-periphery categorical model. Note: dark grey squares represent core nodes; light grey squares represent periphery nodes.

ranked by the continuous model and their associated corene scores.

Determining the variability of subject matter through content analysis on all sixty-one core domains indicated by the categorical model proved cumbersome, therefore only the nodes present in both models were analyzed; this was a limitation with the data capture which could be solved by automated methods in future. Eight nodes from the continuous model were used to highlight the types of websites found within the core and it is important to note that these eight nodes were identified as part of the core structures of both models. The structural position of the core within the network is shown in Figure 1. With the addition of more attributes, the core 61 nodes could be assessed to determine if there were any similarities between them. There may be a correlation between domain membership and content, with certain typologies more likely to exist in the core.

A content analysis of the core eight nodes was conducted to determine if any particular type of website or content was contributing to the core. The top eight nodes all appeared to have similar domain structures (directory sites), while the content varied between the sites. Notably, the website with the highest corene score was offline when the content analysis was conducted two months after data capture. While the continuous model was not selected to represent the network visually, the corene scores demonstrated which core nodes were the most central to the network. The core is represented by the dark grey squares, while the periphery is represented by the light grey squares in Figure 1. Visually, these 61 nodes, given their centralization and large number of ties, represented what would be expected of a core as demonstrated by their position in the network. The large cluster of dark grey squares, shown centre left in Figure 1, represented approximately 60% of the core. Many nodes overlapped, which suggested they held a similar place within the network and may be part of the same community.

4.2. Core-Periphery Analysis. Group densities were also compared to determine the best model to use for the core-periphery analysis and are shown in Table 2. These comparisons helped to identify how connected the core is to itself and the periphery. If the core connections were not above the network average it could have indicated the model was a poor fit for the data. The density measured the proportion of connections that existed as a ratio of all possible connections available within the network [39]. In the categorical model the core density was 0.132, as compared to the overall network density of 0.002. The large discrepancy indicated that there was in fact a core present within the network. The density from the core to the periphery was .019 and from the periphery back to the core it was 0.001. The connection of the core nodes to the peripheral nodes was low, having only formed connections to 1.9% of possible nodes. From periphery to periphery the density was also 0.001, less than the total network. The peripheral connectivity was the same for both the core and other peripheral nodes making connections with only 1.16 nodes. The core nodes connected to an average of eight other members of the core. The core also connected, on average, to approximately 22 periphery nodes. This was in comparison to the periphery which linked to other peripheral nodes at just over one connection per node. These large differences indicated that there was a core present within this network and was consistent with other online communities containing scale-free properties.

In the continuous model the core density was 0.25; meaning 25% of all the possible ties between core members were present. Each core node on average connected to at least two other members in the core. The core to periphery was 0.16, which was also quite large given that the periphery consisted of 1,212 nodes. On average, the core nodes connected to 194 periphery nodes within the network. The periphery back to the core was 0.001 and the periphery to the periphery nodes remained the same at 0.001. While the continuous model had a much denser core, the categorical model retained connectivity while including more nodes. This was ideal for identifying the nodes that possessed a significant number of the connections within the network. Adding the ties from all dyadic pairs within the core, and ties from the core to the periphery in the continuous model, 57% of all the ties in the network were present. In the categorical model this rises to 66%, indicating that the core was responsible for almost two-thirds of all present ties.

4.3. Network Structure. The fundamental assumption underlying a network using websites as nodes was that domains can have social relationships or meaningful connections to other websites [26]. These assumed relationships suggest that websites have properties that can be considered analogous to the connections made by people. Websites at their core

are still operated by users, so while this remains a possible limitation, the hyperlinks are still chosen by individuals, an action no different than choosing friendships offline. Comparing networks across group typologies then should be considered in this context, such as those containing child exploitation materials [4, 30], terrorist networks [40], or abstract structures [25, 32]. Domains hyperlinking to other domains are a product of this decision-making process and do not significantly differ from how edges are formed in other contemporary groups in different contexts. In this study, examining the network characteristics revealed that Tor shared some similarities with these other observed networks. Without an extensive meta-analysis of all network types, it was impossible to place Tor into a specific category.

Comparing results across online networks can help to clarify how online networks operate. For example, a comparison of the cohesion measures indicated that the density of 0.002 found in the Tor network was lower than those found in two regular Internet child exploitation networks (0.05 and 0.11, resp.) [30]. It is possible this observed difference was due to network size; however, networks containing CE materials may be the best comparison available due to criminal domain saturation. Comparing the features of the network to licit alternatives provides little benefit except to simply say that Tor is in fact different. The average distance of 4.95 hops (the sum of all dyadic pairs between nodes) within the Tor network is consistent with the literature regarding online networks exhibiting small-world properties [32]. As defined by Milgram, small-world properties are networks with an average path length of six or less [41]. In addition to containing properties associated with small-world networks the Tor network also had features consistent with a power-law distribution that defines scale-free networks [32]. This was measured through the degree centralization score, which in the Tor network was 42%; this was quite large and surpasses other online networks [32, 42]. Although it was difficult to speculate as to the exact network structure of Tor without a more exhaustive review, the Tor network appeared to share similar properties with other online networks although definitively more centralized.

5. Discussion

5.1. Core-Periphery. The results of the core-periphery model were consistent with the findings regarding the overall degree centralization of the network. As noted above, the reported degree centralization of 42% indicated the network contained properties associated with a scale-free network: few hubs with many connections. This indicated there were some parallel attributes between the regular Internet and the dark web, as both contained features of a scale-free network. A core of 61 nodes (or 5% of the network) was identified and had an overall group density of 0.132. At 18.019, the core to periphery densities were also larger than those of the overall network (0.002). This core contingent accounted for a significant portion of all ties in the network and connections to other core nodes, indicating this data reduction technique may be useful for examining the Tor network. Core-periphery analysis works best in an environment that cannot be divided reasonably into cohesive subgroups [39]. Without a dataset containing attributes to interpret these subgroups or factions, core-periphery allowed for the identification of a subset of nodes containing a significant majority of the connections [39]. The Internet was also typically considered to reflect a core-periphery model as conceptualized by [34, 43] when they describe how links are formed through specific network paths, as opposed to random chance. These authors suggested the Internet was inherently designed in such a way that information travels along designated paths to form a core [34]. The idea that new nodes within the network were more likely to form connections with existing popular nodes. For example, a newly registered domain will link to Google™ at a significantly higher rate than cbc.ca. Again, the Tor network appeared to operate in a similar manner, with websites connected to a small number of central nodes rather than similar peripheral websites. This demonstrated that new nodes, or websites, were more likely to form connections to older, more well-established websites than to newer sites. New domains on Tor were more likely to link to this centralized core than form connections to peripheral nodes. This could be due to the relative anonymity associated with domain creation within Tor, where users are simply unaware of the other nodes' existence. Either way it has implications for law enforcement regarding authorship and domain generation, nodes which connected outside of the norm to more peripheral nodes may be authored or created by the same individual.

The results also deviate from those found in relation to an offline gang network [44], insofar as the current study suggested a much denser core. The similarities between the two types of networks both being elusive and illicit in nature did not seem to link these structures. The core-periphery analysis in the current model is more conducive to Internet-based networks, suggesting that network topography plays a larger role than network typology. Online networks may not be comparable to offline networks despite the symmetries in the content of those networks due to the scale-free nature of online networks. The presence of overlapping nodes indicated they belong to the same community on the network. This finding suggested that connections between websites were not driven by homogeneous characteristics and content. The core-periphery analysis also suggested that connections on the network are dispersed and cannot be tied into reasonably cohesive subgroups. It was also more difficult to determine any key members within the network – obfuscating the entities that produced the majority of illicit activity or content. This made sense given the Tor network is employed by users aiming to avoid detection by law enforcement.

The absence of homophily in the network structures also showed that websites were dissimilar in content and that ties were not created or limited by shared characteristics. However, this can be seen as a benefit for Tor users. Forming connections with dissimilar websites can increase embeddedness and the ability of users to operate without being tracked or monitored, a key objective of Tor users. These ties were also more likely to dissolve, which may contribute to the transient nature of the Tor network [29]. As social disorganization theory suggests, the diversity of users can interfere with the Tor

community's ability to form relationships between users and develop effective communication. Without communication, surveillance of activities on the network is greatly inhibited and dissimilarity between users can instead breed mistrust among the community. Thus, the weak ties and heterogeneity found within network communities lend credence to the possible roles these characteristics have in fostering illicit activity on the Tor network.

Online illicit networks provide an additional advantage for individuals seeking to engage in malicious or illicit activities as compared to offline illicit networks. Unlike offline networks, websites do not need to spend costly resources to maintain social capital and sustain relevancy within the network. Thus, core membership was likely less intensive in online networks, allowing these key sites to maintain their structural positioning within the network. This was different from offline networks where core-periphery structures provided different network contexts. Larger cores allowed the network to sustain its structure if users left, while larger peripheries allowed more adaptability within networks [44]. For law enforcement removing nodes within a network with a large core would have little disruption effect as the pathways through the network are easily attained through other avenues. Targeting of nodes for removal then should not be based upon the principles of disruption or fragmentation but instead should be aimed towards deterrence or desistance where the impact would affect the largest number of users possible. Cryptomarkets and large forums would make ideal targets where users have made investments into accounts whether financially or through gaining trust and reputation. Additionally, dispersed cores are not impacted when peripheral nodes continually enter and exit the network. A single core node does not serve as the only linkage between the cluster of peripheral nodes and the rest of the network. Conversely, removing core nodes within larger peripheries, as in the case of the current study where the core is characterized by scale-free features, will have a substantially larger impact on network traversal. For example, the targeted law enforcement efforts on the "Silk Road" by the FBI saw a large disruption effect for Tor users who were not simply displaced to competing cryptomarkets [12].

In determining if the Tor network shared similar properties to the regular Internet, it is likely that the connections between domains are formed using the same principles [25]. Edge formation was not random and not all nodes within the network have the same probabilities of forming connections [25]. There was, to some degree, a form of preferential selection taking place, where new nodes were more likely to link back to these already established core nodes. The encrypted nature of Tor means information is not routed in exactly the same way, as data still travels through a designated path. Core modeling of the regular Internet further reinforced the use of this analysis in assessing these models within a social network context [34]. The 61 core nodes identified through the categorical model should be compared further with attribute data to speculate about the types of websites comprising this crucial component. The results of the content analysis for the core eight domains indicated that over half are primarily focused on supplying hyperlinks to other domains on Tor. Due to the anonymity and obscurity of Tor, it is likely these directory sites are necessary for users to find content. If the dissemination of information on Tor is reliant upon these directory sites to such a large degree, it has implications for how crime occurs within this context. Future research is necessary to assess if directory sites are linking illegal or criminal websites to other criminal sites or if these sites are forming their own communities.

Following a power-law distribution, the network properties of Tor indicate that it is less likely to recover following domain failures of the hub nodes [25, 43]. Two important considerations are that this leaves the network vulnerable to attack from groups looking to disrupt the network (law enforcement, hacktivists, governments, etc.) or internal causes such as the lack of infrastructure may cause significant disruptions to accessibility and network fluidity. While removing one node remains a challenge and has shown to be a costly endeavour by law enforcement [12], targeting of a select few nodes representing the core structure may have larger and more impactful disruption effects than focusing on the criminal activities within a single domain. The core structure will also regulate how new users can move from domain to domain on Tor likely having to pass through a hub before specific URLs are identified. A hub which is offline for a significant amount of time may vastly restrict the access of users to content which may funnel them into following a different path. Consider the example of a user looking to purchase a firearm in the context of the results. They download the Tor browser and follow the steps and find the Hidden Wiki. This domain has links to a few domains which reportedly sell firearms; however the links are dead or broken. The person decides to try one of the marketplaces and see if someone is selling there. On the marketplace someone may be advertising a different domain which sells firearms because it may go against the policies of the marketplace. The person then follows the hyperlink to the final domain and is able to purchase the gun. Alternatively, in a less centralized network, the domain which sells firearms may be directly hyperlinked by many websites and is accessible by simply clicking it directly from the Hidden Wiki. The addition of one domain does not seem like a preventative factor in stopping a motivated user from purchasing the firearm but due to the factors listed above may be critical. The user needs to find the content on the marketplace, law enforcement has another avenue for intervention and both need to remain stable long enough for this user to make the critical leaps.

6. Conclusions

The Tor network is a relatively unknown and unexplored region of the dark web allowing users to anonymously communicate and browse content through encrypted means. Past research has speculated on the content of Tor through limited studies focused on small sections of the network [5, 7]. A more thorough and systematic process was necessary to conduct analysis of the Tor network. In the current study, an automated tool was used to collect data and social network analysis was applied to examine how websites formed

connections through hyperlinks to other websites. This type of analysis adds a further dimension to how users access content on Tor.

The results of the social network analysis provided insight into the characteristics of the virtual communities formed through hyperlink connections. Support for a homophily effect regarding the way Tor nodes were connecting to others within the network was found. This was also supported by the notion that the Tor nodes had, on average, more incoming ties than the link sites. The core-periphery model identified the main 61 central nodes within the network and guide users through this dark network. Although there was a slight homophily effect regarding the core nodes, websites within the networks appeared to be largely composed of heterophilous connections. As such, similar content did not appear to be the driving force connecting websites. Reference [4] found a similar finding when analyzing online child exploitation communities and suggested that connections with similar content may have a negative impact; similar connected websites may experience decreased traffic to their domains as users can access the information on competitor's sites. When considered under a social disorganization perspective, it is also probable that dissimilar and heterogeneous connections are contributing to the Tor community's inability to effectively surveil deviant users and activity.

Social disorganization principles can provide insight into how network structures may foster illicit activity on Tor. The sparse online networks appear to share similar characteristics to real world crime-prone neighborhoods; they are largely composed of weak and heterophilous ties between domains. Although it was not specifically examined in this current study, Tor is also known for its inherently transient nature; a vast number of domains are created and taken down on a daily basis. Such instability may amplify the facilitation of illicit activity; users specifically choose to operate on Tor and other dark webs to avoid detection and obfuscate their activities. The extension of a criminological theory and concepts should be further explicated to understand how dark web community structures could be fostering illicit activity. The integration of criminology theory is useful to the extent that it can potentially help inform police strategies and reduce criminal activity by targeting formed network communities. Further, it would be advantageous to take a later sample of the data to determine if the characteristics of Tor network structures remain consistent. Due to the sheer volume and size of Tor, measuring change over time within the network structures is needed to refine and corroborate our findings.

As the data was not collected through a randomized sample and may be reflective of the seed websites, the results are not without limitations. For example, if the seed websites were entirely focused on drug markets, without attribute data it would not be possible to tell if the entire sample consists of drug related websites. The seed sites may be biasing the sample based upon content rather than overall network position. The lack of attribute data hampers the ability of the data to be generalizable to a broader sample without first examining what types of websites were collected. Future research should include as many attributes as possible from each website such as content, legality, page views, visitors, registered users, and Bitcoin wallets to allow comparison between the network measures and website characteristics as was previously done in child exploitation research [4]. The attributes would also allow for community detection to be used to examine how these websites were grouping together to form connections. The core-periphery analysis used in the study examined only one set of these communities and it is likely that many others exist. Moreover, applying criminological concepts and measures to the online realm has many challenges and this study did not empirically test social disorganization theory or examine how other structural factors can mediate or influence the relationship between communities and crime [23].

Dark web networks have received significant attention in the past decade, in both the public and political realm, and are being viewed as the only way to protect one's information and privacy in a world where the collection and flow of personal data are nearly inescapable for Internet users. It remains an important task for researchers and law enforcement to pursue a further understanding of how these networks are exploited by individuals and entities for criminal purposes, as this may better inform strategies to disrupt illicit activities on Tor and other dark webs. By understanding the characteristics that foster criminal structures on the network, this could aid enforcement in targeting malicious users and removing illicit content. Still, only a portion of the Tor network was explored in this study, leaving a significant portion still in the dark. Employing social network analysis to examine more domains and the networks they form will allow for a better picture of how users are traversing the Tor network and can better indicate whether social disorganization principles can provide an understanding of crime on Tor.

Appendix

In the seed data collection 1000 domains were found using Google, Reddit, and the Hidden Wiki. Of which 150 were selected using a random number generator to seed the crawler to start the data collection period. This was done to eliminate potential selection bias where if a core structure was identified it would simply represent the seeds chosen. If the core was made up of primarily seed domains the network structure would more accurately represent the ego networks of the seeds rather than a representative sample of the Tor network as a whole. The results of a QAP regression (UCINet 6 v. 6.644) showed that ($R^2 = 0.13$) seed domains did not significantly predict core membership ($\exp(B) = 1.05, p > .05$); however, seed domains did have significantly greater indegree ($\exp(B) = 0.31, p < .001$) and average geodesic distances ($\exp(B) = 0.13, p < .01$). The crawler recursively followed found hyperlinks on each collected webpage subsequently stored in a queue. A webpage was scraped and the XML document which comprises the underlying structure of the webpage was stored in the crawler database. The webpage could then be viewed offline from the database and visually verified by a researcher at a later date for content analysis. Data collection could be terminated by manually cancelling the TDC after a certain time period

or by specifying a criterion. For this study TDC ran until it had collected 1 million webpages and lasted ~90 days or 3 months. The 1 million webpages comprise 1,220 unique domains. Each hyperlink found on a webpage then was collapsed into representing a specific domain. If a hyperlink was found on page 1 or page 820 (the mean number of webpages per website) of Website A both represented a tie to Website B and C and was coded as such. The hyperlinks were then aggregated for each domain and used to generate the networks.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] S. Lu, "What is the dark web and who uses it? The Globe and Mail," https://www.theglobeandmail.com/technology/tech-news/what-is-the-dark-web-and-who-uses-it/article26026082/.

[2] S. Mansfield-Devine, "Darknets," *Computer Fraud & Security*, vol. 12, pp. 4–6, 2009.

[3] S. Dredge, "What is tor? A beginners guide to the privacy tool," *The Guardian*, 2013, http://www.theguardian.com/technology/2013/nov/05/tor-beginners-guide-nsa-browser.

[4] B. G. Westlake and M. Bouchard, "Liking and hyperlinking: Community detection in online child sexual exploitation networks," *Social Science Research*, vol. 59, pp. 23–36, 2016.

[5] C. Guitton, "A review of the available content on Tor hidden services: The case against further development," *Computers in Human Behavior*, vol. 29, no. 6, pp. 2805–2815, 2013.

[6] K. Misata, "The Tor Project," *XRDS: Crossroads, The ACM Magazine for Students*, vol. 20, no. 1, p. 45, 2013.

[7] A. J. Kim, *Community building on the web: Secret strategies for successful online communities*, Addison-Wesley Longman Publishing Co., Inc, Boston, MASS, USA, 2000, http://dl.acm.org.proxy.lib.sfu.ca/citation.cfm.

[8] G. Moura, R. Sadre, and A. Pras, "Bad neighborhoods on the internet," *IEEE Communications Magazine*, vol. 52, no. 7, pp. 132–139, 2014.

[9] N. Christin, "Traveling the Silk Road: A Measurement of a Large Anonymous Online Marketplace," Defense Technical Information Center, 2012.

[10] C. R. Shaw and H. D. McKay, *Juvenile delinquency and urban areas*, University of Chicago Press, Chicago, Ill, USA, 1942.

[11] D. Moore and T. Rid, "Cryptopolitik and the darknet," *Survival*, vol. 58, no. 1, pp. 7–38, 2016.

[12] D. Décary-Hétu and L. Giommoni, "Do police crackdowns disrupt drug cryptomarkets? A longitudinal analysis of the effects of Operation Onymous," *Crime, Law and Social Change*, vol. 67, no. 1, pp. 55–75, 2017.

[13] I. Goodwin, "Book Review: H. Rheingold. 1993. The Virtual Community: Homesteading on the Electronic Frontier. Reading, Massachusetts: Addison-Wesley. ISBN 0-201-60870-7 H. Rheingold. 2000. The Virtual Community: Homesteading on the Electronic Frontier (2nd Edition). C," *Westminster Papers in Communication and Culture*, vol. 1, no. 1, p. 103, 2015.

[14] R. P. Bagozzi and U. M. Dholakia, "Intentional social action in virtual communities," *Journal of Interactive Marketing*, vol. 16, no. 2, pp. 2–21, 2002.

[15] R. J. Sampson and R. B. Groves, "Community structure and crime: testing social disorganization theory," *American Journal of Sociology*, vol. 94, no. 4, pp. 774–802, 1989.

[16] J. Bursik, "Ecological theories of crime and delinquency since Shaw and McKay," *Annals of the American Academy of Political and Social Science*, vol. 338, pp. 119–136, 1984.

[17] R. J. Sampson, "Neighborhood and crime: The structural determinants of personal victimization," *Journal of Research in Crime and Delinquency*, vol. 22, no. 1, pp. 7–40, 1985.

[18] F. E. Markowitz, P. E. Bellair, A. E. Liska, and J. Liu, "Extending social disorganization theory: Modeling the relationships between cohesion, disorder, and fear," *Criminology*, vol. 39, no. 2, pp. 293–319, 2001.

[19] R. Kornhauser, *Social sciences of delinquency*, University of Chicago Press, Chicago, Ill, USA, 1978.

[20] J. Byrne and R. J. Sampson, "Key issues in the social ecology of crime," in *The Social Ecology of Crime*, J. Byrne and R. J. Sampson, Eds., pp. 1–22, Springer-Verlag, New York, NY, USA, 1986.

[21] M. E. Cahill and G. F. Mulligan, "The determinants of crime in Tucson, Arizona," *Urban Geography*, vol. 24, no. 7, pp. 582–610, 2003.

[22] S. Wouter and J. R. Hipp, "A longitudinal test of social disorganizarion theory: Feedback effects among cohesion, social control, and disorder," *Criminology*, vol. 49, no. 3, pp. 833–871, 2011.

[23] C. E. Kubrin and R. Weitzer, "New directions in social disorganization theory," *Journal of Research in Crime and Delinquency*, vol. 40, no. 4, pp. 374–402, 2003.

[24] A. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[25] A. L. Barabási and E. Bonabeau, "Scale-free networks," *Scientific American*, vol. 288, no. 5, pp. 50–59, 2003.

[26] B. Westlake, M. Bouchard, and R. Frank, "Finding the key players in online child exploitation networks," *Policy and Internet*, vol. 3, pp. 1–25, 2011.

[27] M. Bouchard, K. Joffres, and R. Frank, "Preliminary analytical considerations in designing a terrorism and extremism online network extractor," *Intelligent Systems Reference Library*, vol. 53, pp. 171–184, 2014.

[28] B. Monk, R. Allsup, and R. Frank, "LECENing places to hide: Geo-mapping child exploitation material," in *Proceedings of the 13th IEEE International Conference on Intelligence and Security Informatics (ISI'15)*, pp. 73–78, usa, May 2015.

[29] M. McPherson, L. Smith-Lovin, and J. M. Cook, "Birds of a feather: homophily in social networks," *Annual Review of Sociology*, vol. 27, pp. 415–444, 2001.

[30] K. Joffres and M. Bouchard, "Vulnerabilities in online child pornography networks," in *Using Network Analysis to Prevent Crime. Crime Prevention Studies*, A. Malm and G. Bichler, Eds., Criminal Justice Press, New York, NY, USA, 2015.

[31] G. Kadianakis and K. Loesing, "Extrapolating network totals from hidden-service statistics," Tor Tech Report 01(001), 2015, https://research.torproject.org/techreports/extrapolating-hidserv-stats-2015-01-31.pdfRetrieved from.

[32] A.-L. Barabási, "The physics of the Web," *Physics World*, vol. 14, no. 7, pp. 33–38, 2001.

[33] L. Cohen and M. Felson, "Social change and crime rate trends: A routine activity approach," *American Sociological Review*, vol. 44, no. 4, pp. 588–608, 1979.

[34] R. Pastor-Satorras and A. Vespignani, *Evolution and structure of the internet: A statistical physics approach*, Cambridge University Press, Cambridge, UK, 2004.

[35] J. C. Wang and C. C. Chiu, "Recommending trusted online auction sellers using social network analysis," *Expert Systems with Applications*, vol. 34, no. 3, 2008.

[36] S. L. Toral, M. R. Martínez-Torres, and F. Barrero, "Analysis of virtual communities supporting OSS projects using social network analysis," *Information and Software Technology*, vol. 52, no. 3, pp. 296–303, 2010.

[37] M. Bouchard and R. Konarski, "Assessing the core membership of a youth gang from its co-offending network," in *Crime and Networks*, C. Morselli, Ed., Criminology and Justice Series, Routledge, New York, NY, USA, 2014.

[38] H. Jeong, B. Tombor, R. Albert, Z. N. Oltval, and A.-L. Barabásl, "The large-scale organization of metabolic networks," *Nature*, vol. 407, no. 6804, pp. 651–654, 2000.

[39] S. P. Borgatti and M. G. Everett, "Models of core-periphery structures," *Social Networks*, vol. 21, no. 4, pp. 375–395, 2000.

[40] J. Xu and H. Chen, "The topology of dark networks," *Communications of the ACM*, vol. 51, no. 10, pp. 58–65, 2008.

[41] S. Milgram, "The small world problem," *Psychology Today*, vol. 1, no. 1, pp. 61–67, 1967, http://snap.stanford.edu/class/cs224w-readings/milgram67smallworld.pdf.

[42] L. A. Adamic, "The Small World Web," in *Research and Advanced Technology for Digital Libraries*, pp. 443–452, Springer, Berlin, Heidelberg, Germany, 1999.

[43] A. L. Barabási, R. Albert, and H. Jeong, "Scale-free characteristics of random networks: the topology of the world-wide web," *Physica A*, vol. 281, no. 1, pp. 68–77, 2000.

[44] B. Hoppe and C. Reinelt, "Social network analysis and the evaluation of leadership networks," *The Leadership Quarterly*, vol. 21, no. 4, pp. 600–619, 2010.